·方法学研究·

结局测量工具系统评价的报告规范（PRISMA-COSMIN for OMIs 2024）解读

张强[1]，赵文霞[1*]，王子怡[2]，闫乐[1]，梁浩卫[1]，王金钒[3]，方进华[4]，卢存存[5*]

1.450000 河南省郑州市，河南中医药大学第一附属医院
2.730000 甘肃省兰州市，兰州大学公共卫生学院循证社会科学研究中心
3.450046 河南省郑州市，河南中医药大学
4.445300 湖北省恩施土家族苗族自治州，建始县中医医院
5.100700 北京市，中国中医科学院中医临床基础医学研究所
*通信作者：赵文霞，教授；E-mail：wenxia-zhao@163.com
卢存存；E-mail：lu17metrics@163.com

【摘要】 结局测量工具（OMI）是评估患者感受或健康状况的重要工具，但其质量层次不齐。选择健康测量工具的共识标准（COSMIN）为遴选高质量的 OMI 提供了方法学指导，得到广泛认可。近年来，基于 COSMIN 开展的 OMI 系统评价数量激增，然而许多已发表的相关系统评价常未能充分报告关键信息，严重影响了 OMI 系统评价的可重复性和可解释性，从而影响其结果的推广应用。鉴于 PRISMA 2020 未包含报告此类系统评价的所有必要信息，有关学者在前者基础上制定了新的报告规范“PRISMA-COSMIN for OMIs 2024”，以帮助研究人员以清晰、详细和透明的方式撰写和报告 OMI 系统评价。本文结合实例对该规范进行了介绍和解读，旨在帮助国内学者深入理解并有效应用该指南，提高国内 OMI 系统评价的整体质量。

【关键词】 患者报告结局评价；健康结局；结局测量工具；选择健康测量工具的共识标准；系统评价；报告规范

【中图分类号】 R 19 【文献标识码】 A DOI：10.12114/j.issn.1007-9572.2024.0465

Interpretation of the Reporting Guideline for Systematic Reviews of Outcome Measurement Instruments (PRISMA-COSMIN for OMIs 2024)

ZHANG Qiang[1], *ZHAO Wenxia*[1*], *WANG Ziyi*[2], *YAN Le*[1], *LIANG Haowei*[1], *WANG Jinfan*[3], *FANG Jinhua*[4], *LU Cuncun*[5*]

1.The First Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou 450000, China
2.Evidence-Based Social Science Research Center, School of Public Health, Lanzhou University, Lanzhou 730000, China
3.Henan University of Chinese Medicine, Zhengzhou 450046, China
4.Jianshi County Chinese Medicine Hospital, Enshi 445300, China
5.Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China
**Corresponding authors: ZHAO Wenxia, Professor; E-mail: wenxia-zhao@163.com*
LU Cuncun; E-mail: lu17metrics@163.com

【Abstract】 Outcome Measurement Instruments（OMI）are essential tools for evaluating patients' experiences or health status, but their quality levels often vary. The COnsensus-based Standards for the selection of health Measurement INstruments（COSMIN）provide methodological guidance for selecting high-quality OMIs, which is widely recognized. In recent

基金项目：河南省中医药科学研究专项课题（2024ZY2046）；河南省中医学“双一流”创建科学研究专项（STG-ZYX02-202117、HSRP-DFCTCM-2023-4-02）；河南省科技研发计划联合基金（优势学科培育类）项目（242301420112）

引用本文：张强，赵文霞，王子怡，等.结局测量工具系统评价的报告规范（PRISMA-COSMIN for OMIs 2024）解读［J］.中国全科医学，2025. DOI：10.12114/j.issn.1007-9572.2024.0465.［Epub ahead of print］［www.chinagp.net］

ZHANG Q, ZHAO W X, WANG Z Y, et al. Interpretation of the reporting guideline for systematic reviews of outcome measurement instruments (PRISMA-COSMIN for OMIs 2024)［J］. Chinese General Practice, 2025.［Epub ahead of print］

©Editorial Office of Chinese General Practice. This is an open access article under the CC BY-NC-ND 4.0 license.

years, there has been a sharp increase in the number of systematic reviews of OMI based on COSMIN. However, many published related systematic reviews often fail to fully report key information, significantly impacting the reproducibility and interpretability of such systematic reviews, thus affecting the dissemination and application of their results. As the PRISMA 2020 does not include all necessary information for reporting such systematic reviews, the scholars have developed a new reporting standard, "PRISMA-COSMIN for OMIs 2024", to assist researchers in writing and reporting systematic reviews of OMI in a clear, detailed, and transparent manner. This article introduces and interprets this guideline through an example paper, aiming to help domestic scholars better understand and effectively apply the guideline, thereby enhancing the overall quality of systematic reviews of OMI conducted in China.

【Key words】 Patient reported outcome measures; Health outcome; Outcome measurement instrument; COnsensus-based Standards for the selection of health Measurement INstruments; Systematic review; Reporting guideline

结局测量工具（outcome measurement instrument，OMI）是评估患者主观感受或健康状况的重要工具[1-2]。选择和使用可靠的 OMI 对于明确患者的结局改善情况和临床干预的效果至关重要，然而，针对同一结局的多个测量工具的质量常参差不齐。选择健康测量工具的共识标准（COnsensus-based Standards for the selection of health Measurement INstruments，COSMIN）是目前认可度较高且被广泛用于遴选高质量 OMI 的方法学指南[3-4]。系统评价作为一种方便、低成本和受欢迎的研究方法可用于整合不同 OMI 的研究，有助于比较和整合不同研究的结果[5]。近年来，基于 COSMIN 指南开展的 OMI 系统评价数量显著增加[6-8]，然而，在已发表的报告中经常缺少关键信息（例如，工具的可行性解释、数据的整合方法和测量属性的原始数据等），这种不完整的报告严重影响系统评价结果的重现和用户选择适合测量工具的判断[8-9]。尽管 PRISMA 2020 是专门用于报告系统评价的首选报告规范，但其主要侧重于干预措施相关内容的报告[10]，而 OMI 系统评价则需要对每个测量属性进行评价和报告。另外，OMI 系统评价与干预性系统评价二者在效应的测量和证据合成的方法上也存在差异，因此，PRISMA 2020 的条目并不完全适用于 OMI 系统评价[8, 11]。鉴于此，有关学者在 PRISMA 2020 基础上制定并于近期在 Journal of Clinical Epidemiology 发表了专门用于指导 OMI 系统评价报告的报告规范——PRISMA-COSMIN for OMIs 2024[11-12]。本文结合实例对其报告清单进行解读，以期帮助国内学者更好地理解和应用该规范，从而为国内学者制作或更新 OMI 系统评价提供参考。

1 PRISMA-COSMIN for OMIs 2024 制定过程简介

PRISMA-COSMIN for OMIs 2024 规范是在 PRISMA 2020 的框架基础上进行修改、补充和删减而成[11]。制定过程共包括 7 个步骤。第一步是项目启动：研究团队在 EQUATOR 网站注册了该规范的制定过程，并在学术期刊上发表了制定方案[8]，同时成立了指导委员会和技术咨询小组。第二步是准备阶段：通过检索文献确定潜在适用的条目，形成初始条目列表并将其应用到三种不同类型的系统评价进行修订补充，结果经指导委员会和技术顾问小组确定后形成初始条目清单。同时邀请参与后续德尔菲调查的潜在小组成员，他们具有设计、制作、发表或应用 OMI 系统评价的背景。第三步是德尔菲调查：通过三轮德尔菲调查对清单中推荐的 49 个初始条目进行确认。103 人完成了第一轮调查，他们采用"非常不同意""不同意""中立""同意""非常同意"五分制李克特量表对于每项条目是否应该纳入，以及措辞是否清楚进行了评价，同时允许提出补充条目。最终 13 个条目达成了共识，删除了 4 个条目，新增了 1 个条目。78 人完成了第二轮调查，对包括第一轮未达成共识的条目和新增条目进行了确认，最终 19 个条目达成共识，删除了 4 个条目。78 人完成了第三轮调查，对包括第一轮或第二轮没有达成共识的条目以及修改措辞的条目进行了确认，最终 12 个条目达成了共识。第四步是工作会议：最后通过 24 人的混合工作组会议对第三轮调查中没有达成共识和措辞修改的条目进行确认，对每个条目投票（"纳入""排除""弃权"）。超过 70% 的纳入或排除即认为达成共识，最终又对另外 10 个条目达成了共识，并删除了 1 个条目。此外还讨论了解释性文件和预测试及宣传的情况。第五步是起草报告规范：写作者和审阅者团队起草了规范，编辑团队创建了报告清单、解释性文件和流程图。第六步是预测试：65 位作者测试了报告规范，并对条目措辞提出了小的建议。第七步是项目结束会议：执行委员会确定了最终版本规范，包括 54 个（子）条目的全文报告清单与 13 个条目的摘要报告清单、1 个解释性文件和流程图[11-12]。

2 PRISMA-COSMIN for OMIs 2024 条目解读

PRISMA-CONMIN for OMIs 2024 规范清单[11-12]包括标题、摘要、通俗语言摘要、开放科学、前言、方法、

结果和结论等部分，共包含 54 个（子）条目，见表 1，其中 13 个为摘要报告条目。鉴于 OMI 系统评价方法目前在护理学领域应用较多，故本文选择一篇 2023 年发表于《中华护理杂志》题为“基于 COSMIN 指南对糖尿病患者健康素养评估工具的系统评价”的实例对该报告规范的条目进行解读[13]。

3 PRISMA-CONMIN for OMIs 2024 清单的解

表 1 PRISMA-CONMIN for OMIs 2024 清单[11]
Table 1 Checklist for PRISMA-CONMIN for OMIs 2024

章节	主题	条目	内容	位置
标题				
	标题	1	识别为一项系统评价，并包括（顺序不限）：关注的结局领域、人群、结局测量工具名称 / 类型以及测量属性	
摘要				
	开放科学			
	基金	2.2	说明系统评价的主要资金来源	
	注册	2.3	提供注册网站名称和注册号	
	背景			
	目的	2.4	说明该系统评价的主要目的或所要解决的问题	
	方法			
	合格标准	2.5	说明系统评价的纳入和排除标准	
	信息来源	2.6	说明用于确定研究的信息来源（如数据库、注册网站）以及每个来源最后检索的日期	
	偏倚风险	2.7	说明用于评估纳入研究偏倚风险的方法	
	测量属性	2.8	说明用于评估测量属性结果的方法	
	合成方法	2.9	说明用于呈现和整合结果的方法	
	结果			
	纳入的研究	2.10	提供纳入的结局测量工具和研究报告的总数	
	结果的整合	2.11	呈现合成的结局测量工具结果，同时报告证据的确定性	
	讨论			
	证据的局限性	2.12	提供系统评价所纳入证据的局限性的简要总结，如偏倚风险、一致性和不精确性	
	解释	2.13	提供对结果和重要意义的一般性解释	
通俗语言摘要				
	通俗语言摘要	3	如果期刊允许的话，请提供一份包括背景信息和主要发现的通俗语言摘要	
	开放科学			
	注册和方案	4a	提供系统评价的注册信息，包括注册网站名称和注册号，或说明该系统评价未进行注册	
		4b	说明访问系统评价方案的途径，或说明没有准备系统评价方案	
		4c	描述并解释对注册或方案中的信息的任何修改	
	资助	5	描述系统评价中的资金或非资金支持来源，并说明资助者在系统评价中的作用	
	利益冲突	6	声明系统评价作者的任何利益冲突	
	数据、代码和其他材料的可获取性	7	报告下面哪些材料是公开可获取的以及访问途径，如数据收集表模板，从纳入研究中提取的数据，用于所有分析的数据，分析代码，以及其他所使用的材料	
前言				
	原理	8	描述在现有知识背景下开展系统评价的依据	
	目的	9	说明系统评价的目的或所要解决的问题，包括关注的结局领域、人群、结局测量工具名称 / 类型，以及测量属性	
方法				
	遵循的指南	10	说明用于开展系统评价的方法和 / 或指南，并提供相应的参考文献	
	合格标准	11	说明系统评价的纳入和排除标准	
	信息来源	12	说明用以确定研究的所有数据库、注册网站、预印本服务器、网站、组织、参考文献列表及其他检索或咨询的来源，同时说明每个来源最后一次检索或咨询的日期	
	检索策略	13	提供所有数据库、注册网站和其他网站的完整检索策略，包括所使用的滤器和限制	
	筛选过程	14	说明用于决定一项研究是否符合系统评价纳入标准的方法，如包括每个检索记录和获取的报告由多少位研究者筛选，是否独立进行，以及（若适用）过程中使用的自动化工具 / 人工智能的详细信息	



（续表 1）

章节	主题	条目	内容	位置
	数据收集过程	15	说明从报告中收集数据的方法，如包括每份报告由多少位研究者收集数据，是否独立进行，从原始研究作者处获取或确认数据的任何过程，以及（若适用）过程中使用的自动化工具 / 人工智能的详细信息	
	数据资料	16	列出并定义提取的数据内容（如研究人群和结局测量工具的特征、测量属性结果以及可行性和可解释性）。描述处理任何缺失或不明确信息的方法	
	研究的偏倚风险评估	17	说明评估纳入研究偏倚风险的方法，如包括使用的工具细节，每个研究由多少位研究者进行评估，是否独立进行，以及（若适用）过程中使用的自动化工具 / 人工智能的详细信息	
	测量属性	18	说明用于对单个研究以及汇总或合并结果的测量属性结果进行评估的方法，如包括有多少位研究者对每项研究进行评估以及是否独立进行	
	合成方法	19a	描述用于确定哪些研究符合纳入每个整合分析的过程	
		19b	描述用于整合结果的任何方法	
		19c	若适用，描述用于发现研究结果间不一致的可能原因的任何方法（如亚组分析）	
		19d	若适用，描述用以评估合成结果稳健性所进行的任何敏感性分析	
	确定性评估	20	描述用于评估证据体确定性（或可信度）的任何方法	
	制定建议	21	若适用，描述用于制定针对特定用途的结局测量指标的适用性的建议的任何方法	
结果				
	研究筛选	22a	描述检索和筛选过程的结果，从检索识别到的记录数量到纳入系统评价的研究报告数量，最好使用流程图展示。若适用，还需报告最终纳入的结局测量工具数量以及每个结局测量工具相关的研究报告数量	
		22b	引用可能符合纳入标准但被排除的研究报告，并解释他们被排除的原因	
	结局测量工具的特征	23a	呈现每个纳入的结局测量工具的特征，并提供参考文献	
		23b	若适用，呈现每个纳入的结局测量工具的可解释性	
		23c	若适用，呈现每个纳入的结局测量工具的可行性	
	研究特征	24	引用纳入的每项评估一个或多个测量属性的研究报告，并呈现其特征	
	研究的偏倚风险	25	呈现每项纳入研究偏倚风险评估的结果	
	单个研究的结果	26	针对所有测量属性，每项研究呈现：（a）报告的结果；（b）基于质量标准的评估结果，最好使用结构化的表格或图形	
	合成的结果	27a	呈现所有整合的结果。对于一个结局测量工具的每个测量属性，呈现：（a）汇总或合并的结果；（b）基于质量标准的总体评估结果	
		27b	若适用，呈现对研究结果之间可能存在的不一致性原因的所有调查结果	
		27c	若适用，呈现评估整合结果稳健性的所有敏感性分析结果	
	证据确定性	28	呈现对每个评估的结局测量工具的每个测量属性的证据体确定性（或可信度）评估的结果	
	建议	29	若适用，对特定用途的合适结局测量工具提出建议	
讨论				
	讨论	30a	在其他证据的背景下，对结果进行一般性的解释	
		30b	讨论系统评价中纳入证据的局限性	
		30c	讨论系统评价过程中存在的局限性	
		30d	讨论结果对实践、政策和未来研究的意义	

注：PRISMA-COSMIN for OMIs 2024 摘要报告清单的条目 2.1 与全文报告清单的条目 1 相同。

读

考虑到条目数量及具体内容较多，故不再在此重复列出，见表 1，此处仅给出条目解释与实例论文的报告情况。

3.1 标题

3.1.1 条目 1：题目中应包含“系统评价（systematic review）”术语，从而有助于数据库的恰当标引，也便于患者、卫生保健提供者、政策制定者和研究人员等发现和检索。考虑到“文献综述（literature review）”和“证据合成（evidence synthesis）”等无法区分是否“系统”，故不推荐使用[11]。结局领域、人群、结局测量工具名称 / 类型以及测量属性是确定系统评价的研究目的或问题的 4 个关键要素，因此，有必要在题目中报告这些信息。实例以“基于 COSMIN 指南对糖尿病患者健康素养评估工具的系统评价”为题目，很好地体现了该条目的要求。

3.2 摘要

3.2.1 开放科学

3.2.1.1 条目 2.2：除非期刊另有指定报告位置，否

则应在摘要中说明开展系统评价的主要资助来源。对于会议摘要，也须报告此信息。实例在其发表期刊的专有位置报告了得到“浙江省医药卫生科技计划项目（2023KY134）”资助。

3.2.1.2 条目 2.3：系统评价的注册记录能够提醒其他研究人员有正在进行的系统评价，作为公共记录，注册有助于识别未发表的系统评价，减少重复工作和报告偏倚所导致的研究浪费[14]。除非期刊另有指定报告位置，否则摘要中应记录系统评价注册的数据库名称和注册号。会议摘要也应报告这些信息。实例研究未提及注册的相关情况。

3.2.2 背景

3.2.2.1 条目 2.4：摘要中应简洁表述系统评价的目的或研究问题，其内容需涵盖结局领域、人群、OMI 名称或类型以及测量属性这 4 项关键要素。实例报告其目的是系统评价糖尿病患者健康素养评估工具的方法学质量与测量属性质量，从而为相关人员选择最佳评估工具提供依据。

3.2.3 方法

3.2.3.1 条目 2.5：合格标准用于判定研究是否符合纳入评价条件的主要特征，可帮助读者理解系统评价的范围并检查纳入决策。与研究目的类似，合格标准通常也涉及结局领域、人群、OMI 的名称或类型、测量属性 4 个关键要素。纳入研究通常应评估一个或多个关注的测量属性、OMI 的开发，或报告其可解释性和可行性。实例未在摘要中描述纳入和排除标准。

3.2.3.2 条目 2.6：作者应在摘要中简要描述检索或咨询的信息来源（如数据库和注册网站），并注明各信息来源的最后检索日期，以便读者评估系统评价的完整性与时效性。若使用了多个信息来源，应报告数据库的总数。实例摘要中报告检索了 PubMed、Embase、CINAHL、Web of Science、中国知网、维普网、万方数据知识服务平台及中国生物医学文献数据库，并给出了具体检索时限。

3.2.3.3 条目 2.7：偏倚风险指研究因设计、实施或分析缺陷而系统性偏离事实的可能性。系统评价通过评估偏倚风险判断研究的有效性，因此作者应描述评估偏倚风险所用的工具和方法。若遵循 COSMIN 或 OMERACT[15]指南，可将条目 2.7、2.8 和 2.9 进行汇总报告。实例摘要中报告了使用 COSMIN 系统评价指南评价纳入的工具。

3.2.3.4 条目 2.8：为便于解释结果，作者需要明确评估测量属性结果的评估方法，包括单个研究或汇总研究中测量属性的评估方法。实例摘要中报告了使用 COSMIN 系统评价指南评价纳入的工具，并形成最终推荐。

3.2.3.5 条目 2.9：多个研究报告的测量属性结果通常通过定性归纳总结进行整合。整合结果的具体方法应在摘要中进行说明。如果字数允许，也可以提供关于证据确定性或信心的评估方法。实例摘要中报告了使用 COSMIN 系统评价指南评价纳入的工具，但未明确描述测量属性结果的具体整合方法。

3.2.4 结果

3.2.4.1 条目 2.10：作者应提供纳入的 OMI 的总数及研究报告的数量，以便读者了解系统评价中证据的范围。若发现同一 OMI 的不同版本，也应予以报告，因每个版本的 OMI 均被视为独立工具，但不同语言版本除外。实例摘要中报告共纳入 15 篇文献，包含 12 种糖尿病患者健康素养评估工具，但未报告是否存在不同版本的工具。

3.2.4.2 条目 2.11：摘要中应报告主要整合结果，即与评价目标最为相关的结果。若一项研究评估了所有测量属性，且预先设定内容效度和结构效度对结论至关重要，则至少应提供最相关的 OMI 测量属性的整合结果。在提供整合结果的同时，若篇幅允许，可一并提供每项整合结果的证据确定性或信心评级结果。实例摘要中报告了纳入评估工具的测量属性汇总结果，并提供了基于 GRADE 方法形成的推荐。

3.2.5 讨论

3.2.5.1 条目 2.12：摘要应简要描述研究局限性，如不完整性、不一致性和不精确性等，从而有助于读者正确解读或使用研究结果决策。实例未在摘要中报告局限性。

3.2.5.2 条目 2.13：为帮助读者理解结果，作者应综合归纳主要发现，涵盖结论明确性、重要不确定性及需进一步研究的领域。若缺乏高质量证据，应明确告知读者。若当前系统评价与以往研究存在明显差异，应予以解释。作者还应结合系统评价证据阐述研究对政策或实践的潜在意义。实例摘要的结论部分给出了两个评估工具的推荐及理由，但未说明其他情况。

3.3 通俗语言摘要

3.3.1 条目 3：在期刊允许的情况下，作者应提供一份通俗语言摘要，即科普摘要，即使用通俗易懂的非技术语言描述主要发现，这有助于患者非专业人员的理解和应用，从而扩大其研究结果的传播范围和影响力[16]。如果期刊不允许简单语言摘要作为常规摘要的一部分，可以将其作为补充材料提供。实例未提供通俗语言摘要，这可能与期刊要求有关。

3.4 开放科学

3.4.1 条目 4a：作者应报告系统评价的注册网站及注册号或 DOI 号，这样有助于研究被检索和发现；同时也便于读者比较注册方案与全文研究之间的差异，判断系统评价结果的可靠性。实例未提供注册的相关信息。

3.4.2 条目 4b：作者应说明系统评价方案的获取方法。

如果作者未准备方案或拒绝他人访问，也应说明。实例未提供研究方案的相关信息。

3.4.3　条目 4c：在早期仔细考虑评价和分析方法，可以减少方案制定后的不必要变化。但是，前期常可能无法考虑到后期的所有情况，此时需要对方案进行修订。为提高研究的透明度，作者应报告所有修改的细节及原因，即使没有修订也应报告。实例未提供研究方案的修订信息。

3.4.4　条目 5：作者应明确说明系统评价受到的资助来源，包括研究人员的资助、商用数据库访问权限等。如果资助者参与了系统评价的某一环节，应描述资助者的参与方式，说明可能引入的潜在偏倚及其对研究结果的影响[17-18]。实例报告其基金项目为“浙江省医药卫生科技计划项目（2023KY134）”，但未提供每位作者的个人资助信息。

3.4.5　条目 6：众多研究已经表明利益冲突能够影响科学研究结果的公正性和可靠性，系统评价亦不例外，且系统评价的结果常被用于制定临床实践指南，影响更大[17-18]，因此，应该报告是否存在利益冲突，如果存在，应该声明如何处理这些冲突。实例报告“所有作者均声明无利益冲突”。

3.4.6　条目 7：数据、分析代码和其他资料的公开有助于其他研究者检查数据是否准确、重现研究结果或了解分析方法。因此，作者应该报告数据提取表模板、提取的数据、用于分析的数据及分析代码等。这些资料可作为附件提交，也可存放在公开数据库，但应提供获取方法。实例未报告该条目要求的信息。

3.5　前言

3.5.1　条目 8：OMI 系统评价的目的主要包括选择最适合特定用途的 OMI，或填补有关 OMI 测量属性的知识空白。详细说明理由与依据有助于读者理解开展该系统评价的必要性及其对现有知识的贡献。特别是如果已有大量类似的系统评价，作者应解释开展当前研究的必要性。实例报告虽然糖尿病患者健康素养测评工具较多，但测量内容和评价方法存在差异，且缺少对此类工具的比较研究，不利于选择合适的测量工具。

3.5.2　条目 9：简明扼要地表述 OMI 系统评价的研究目的或问题，有助于读者理解系统评价的范围以及方法是否足以解决问题。目标或问题应涵盖四个关键要素，即结局领域、人群、OMI 名称或类型，以及测量属性。也可将研究目的或问题与研究依据联系起来。实例报告其目的旨在筛选出最佳的糖尿病患者健康素养测评工具，为准确有效地开展评价提供循证医学证据。

3.6　方法

3.6.1　条目 10：目前有多种方法学指南可用于指导 OMI 系统评价的开展，如 COSMIN 系统评价指南、OMERACT 滤器以及 JBI 证据合成手册等[3, 15, 19]。明确地报告所使用的方法学或参考的指南，并具体说明或引用使用的版本和清单，有助于读者判断研究是否遵循了既定指导原则并采用了高质量的方法。实例未报告相关信息。

3.6.2　条目 11：纳入和排除标准的制定同样涉及结局领域、研究人群、OMI 的名称或类型，以及测量属性 4 个关键要素。建议定义这些要素时，应明确使用的词典或分类法以避免术语之间的差异[11]。此外，纳入和排除标准还可能涉及研究的语言和出版状态，如是否纳入非英文文献以及会议摘要。实例纳入标准报告了纳入糖尿病患者，研究内容为评价健康素养量表测量属性，研究类型为评估工具的编制和验证等，同时排除标准中报告排除非中英文文献以及综述等，但未报告是否排除会议摘要。

3.6.3　条目 12：作者应详细描述所有信息来源，包括数据库、注册网站、有关组织和参考资料的名称（及链接）和检索日期。可参照 PRISMA-S 规范报告文献的检索过程[20]。实例报告采用主题词结合自由词方法检索了 PubMed、Embase、CINAHL、Web of Science、CNKI、维普网、万方数据知识服务平台及中国生物医学文献数据库以及具体的检索时限。

3.6.4　条目 13：为增强透明度和可复现性，需报告所有数据库的详细检索策略。不同数据库方法有异，仅报告单一数据库的检索策略或许会影响其他数据库的评估。由于期刊篇幅限制，详细检索策略可作为附件提交。此外，可运用谷歌学术等进行补充检索，同时可利用滤器提高检索效率。实例提供了中、英文检索词，但只有 PubMed 数据库的完整检索策略。

3.6.5　条目 14：文献筛选通常先依据标题和摘要初筛潜在合格研究，随后进行全文审查，必要时联系研究人员确认。可采用混合筛选方法（如自动化或人工智能技术）进行排除或优先性排序。应报告筛选人员数量、是否独立筛选以及处理分歧的步骤。作者需详细描述纳入决策过程，以帮助读者评估筛选过程中的潜在错误。实例报告由两名经过方法学培训的研究者独立筛选文献，分歧通过与第三名研究者讨论解决。先阅读题目与摘要进行初筛，再进一步阅读全文以确定是否最终纳入。

3.6.6　条目 15：应报告从纳入报告中获取数据的方法，以便读者评估数据潜在错误。若使用软件从图片提取数据，需详细说明所用软件及其版本；若使用已发表的数据提取表（如 COSMIN 指南中的数据提取表），应引用出处。实例仅报告由两名经过方法学培训的研究者独立整合资料，分歧通过与第三名研究者讨论解决，但未报告获取数据的详细方法。

3.6.7　条目 16：作者应于方法学部分报告从各纳入研

究中提取的数据及信息，以便读者理解其内容，并为类似研究的数据收集提供参考。若存在重要信息缺失，可联系原始研究作者获取，或对缺失信息进行合理假设，且这些信息均应详细报告。实例未在方法学部分报告提取数据的具体内容。

3.6.8　条目 17：偏倚可影响甚至完全扭曲研究结果，评估偏倚风险是制作系统评价的关键步骤[21]。此外，用户需了解纳入研究的偏倚风险，以帮助合理解释和评估证据可信度。现有多种工具可用于评估结局指标测量特性研究的局限性，如 COSMIN 偏倚风险清单[4]和最小临床重要改变可信度评估工具[22-23]。作者应报告所用偏倚风险工具的版本及评价标准，说明任何调整，以便读者评估工具适用性。同时应说明参与评估的研究者人数、是否独立评估及分歧解决方法。若使用自动化或人工智能工具评估偏倚风险，需说明使用方法。此外，还应报告对每个测量属性做出的偏倚判断的总体风险。实例报告使用 COSMIN 偏倚风险清单评价了纳入工具测量特性的方法学质量，并采用“很好”“良好”“模糊”“不良”4 级评分方法，对某一特性的总体风险采用最低计分原则。

3.6.9　条目 18：为便于解释结果，作者需详述单个研究及不同研究中用于评价测量属性结果的标准。若对构念效度和反应度进行了评价，应明确用于评估这些测量属性结果的先验假设，如关注的 OMI 与比较的 OMI 之间的预期方向和大小以及相关组间的得分差异[12]。在报告评价方法时，应说明是由一位或两位研究者评分、是否使用参考手册，以及独立评价时如何解决分歧，以便读者识别潜在错误。实例报告使用 COSMIN 质量标准评价了量表的 9 个测量属性，分为“充分”“不充分”和“不确定”三个级别，但未涉及评价人员的相关内容。

3.6.10　条目 19a：在整合相关 OMI 的测量属性前，需确定哪些研究适合纳入各项合成分析。通常是对报告同一结局指标相同测量属性的多项研究结果进行整合。鉴于研究结果的不一致性或不同人群差异可能影响整合研究的选择决策，且这些决策常带有主观判断并可能改变整合结果，故应报告选择过程及依据，以确保决策透明。实例报告了对量表的某一测量属性在不同研究间的评价一致，其总体结果为充分、不充分和不确定；若不一致且无法解释原因，则总体结果为不一致。

3.6.11　条目 19b：在 OMI 系统评价中，最常见的合成方法为对内容效度、构念效度、跨文化效度 / 测量不变性等测量属性结果进行定性总结。作者应报告已证实和未证实假设的数量，或单项研究中测量属性结果的范围。对于部分测量属性（如内部一致性、可靠性、测量误差、构念效度和反应性），虽可利用 Meta 分析定量整合，但在 OMI 系统评价中不常见，因为通常不采用这些结果的点估计值[11]。然而，无论选用何种合成方法，作者均应提供充分的信息。实例未在方法学部分报告测量属性的整合方法。

3.6.12　条目 19c：若作者探索了各研究结果间不一致的原因，应详细说明所探索的具体原因及采用的方法。可能的原因有参与者或结局指标的特征、研究的偏倚风险、方法及时效性等。若上述原因可解释结果间的不一致，可进行亚组分析，即将研究分为亚组并比较不同亚组的结果。实例未在方法学中提及探索不同研究结果间不一致的原因和方法。

3.6.13　条目 19d：敏感性分析用于检验系统评价过程中不同决策对结果稳健性的影响。若作者进行了敏感性分析，应提供详细内容（如剔除高偏倚风险研究或采用其他整合方法等），以便读者评估其适当性并重现结果。理想状态下，敏感性分析应预先设定，但若在评价过程中出现不可预测问题，也需进行。实例未在方法学中提及敏感性分析。

3.6.14　条目 20：通常作者运用特定标准评估每个 OMI 测量属性证据的确定性。所考虑因素包括研究设计局限性、各研究结果一致性、样本量（不精确性）及研究对拟解决问题的直接程度。可采用系统明确的方法（如改良的 GRADE 方法[24]）评估这些因素，作者应报告评估整合结果时考虑的因素与标准，以便读者了解哪些因素纳入确定性评估，评估潜在错误并便于结果重现。实例报告使用改良的 GRADE 方法进行证据质量分级，主要考虑偏倚风险、不一致性、间接性、不精确性 4 个降级因素，最终每个测量属性可分为高、中、低、极低 4 个级别。

3.6.15　条目 21：OMI 系统评价的主要目的在于筛选出最适合特定用途的 OMI，或识别 OMI 测量属性中的知识空白。若目的是选择最适合的 OMI，作者应提出适用性建议，并详述方法与过程。若资助方不允许提出建议或与系统评价要求不符，则可不提出建议。实例在方法学报告了形成测量工具推荐的方法。

3.7　结果

3.7.1　条目 22a：作者最好以流程图报告检索及各筛选步骤的详细结果，以便读者了解检索过程并在必要时重现结果。若为更新先前的系统评价，应报告当前的检索与筛选过程，并明确先前系统评价中所包含的研究报告及 OMI 数量。除报告纳入的研究报告数量外，作者还应报告纳入的 OMI 数量以及每个 OMI 对应的研究报告数量。实例报告了去重阶段、题目与摘要筛选阶段、全文筛选阶段排除的文献数量，但未采用流程图，也未报告 OMI 与研究报告的对应数量。

3.7.2　条目 22b：识别被排除的研究记录有助于读者评估系统评价的有效性和适用性。应提供一份被排除研究

报告的列表清单，其中列出参考文献及排除原因，包括那些看似符合纳入标准但因人群不符合要求等原因被排除的研究，以及无法获取全文或不符合语言要求的报告等。此信息可在报告正文或补充材料中列表呈现。实例未提供排除研究的列表清单。

3.7.3 条目 23a：报告纳入系统评价的 OMI 的详细特征有助于读者了解所纳入 OMI 情况及系统评价的适用性。特征信息主要包括关注的结局领域、开发的目标人群、使用方式、回忆期、（子）量表及条目数量、回答选项、分数范围或评分方法、开发的原始语言及可用的翻译版本等[11]，若有其他适用特征也可一并报告。实例在正文中使用表格和文字描述了纳入文献的作者、发表年份、国家、样本量、条目 / 维度数、重测时间以及评估工具的名称等信息。

3.7.4 条目 23b：报告 OMI 的可解释性信息有助于评估其适用性并为选择最合适的 OMI 提供依据，在以筛选特定用途 OMI 为目的的系统评价中尤为关键。应提供诸如评分分布、缺失条目比例、下限和上限效应以及最小重要改变等信息[11]，同时也可报告在预期使用情境下从 OMI 评分解释中得出的含义及信心程度。这些内容可以表格形式在正文或作为补充材料呈现，以便于比较不同 OMI 的特征。实例未报告纳入 OMI 的可解释性信息。

3.7.5 条目 23c：报告 OMI 的可行性信息有助于评估其适用性及应用的容易程度，在为特定用途选择最合适的 OMI 时尤为重要。可行性信息涵盖使用类型及容易程度、OMI 长度、完成时间、所需患者能力、标准化及评分计算的难易程度、版权、成本、所需设备以及审批要求等[12]。这些内容可通过表格形式在正文或补充材料中呈现，以便于比较不同 OMI 的特征及检索相关资料。实例未报告纳入 OMI 的可行性信息。

3.7.6 条目 24：报告纳入研究的细节信息，包括人群特征（样本量、年龄、性别等）、OMI 管理特征（国家、语言等）、响应率、测量属性、资助来源及利益冲突等[12]，有助于读者理解系统评价的适用性。可在正文或附件中以表格或图形形式呈现每项研究的关键特征，并提供对应的参考文献。实例在正文中使用表格和文字描述了纳入文献的作者、发表年份、国家、样本量、条目 / 维度数、重测时间以及评估工具名称等信息，并提供了相应参考文献。

3.7.7 条目 25：为便于读者评判纳入研究结果的可信度，作者应报告每项纳入研究的偏倚风险评价结果。以表格或图形呈现各研究在不同测量属性上的偏倚风险为佳，此信息可在正文或附件中提供。展示偏倚风险评级时，可与研究提取结果及测量属性的质量标准评级相结合，以便读者全面理解研究结果及其可信度。实例以表格形式呈现了纳入量表的方法学质量与测量属性质量评价结果。

3.7.8 条目 26：呈现单个研究在测量属性方面的结果，有助于理解其对每个 OMI 结论的贡献，且便于他人进行二次分析或更新评价。可在正文或附件中以表格、图形或森林图等形式呈现相关结果，具体呈现方式取决于测量属性特点。不同研究中关于同一 OMI 的相同测量属性结果应归类呈现，并依据预定义质量标准进行评级。实例在正文中以表格和文字形式报告了纳入研究中测量工具的测量属性结果。

3.7.9 条目 27a：系统评价的用户依赖于所有报告的综合结果，以便基于完整且无偏倚的证据进行决策。选择性报告结果属于一种偏倚，所以需透明地报告所有结果。在对测量属性进行评估的系统评价中，结果通常以定性总结的方式呈现，有时也通过统计方法（如 Meta 分析）得出。应同时提供汇总结果以及基于质量标准的总体评级结果。对于多维度的 OMI，需分别报告所有子量表的汇总结果和评级结果。实例在正文中以表格和文字形式报告了纳入 OMI 的测量属性汇总结果及 GRADE 评价结果。

3.7.10 条目 27b：报告所有对研究结果不一致的可能原因的调查结果，对系统评价的用户至关重要。理解影响结果变异性的因素有助于用户做出决策。此外，展示所有研究结果对未来研究的设计也意义重大，因为这些结果可帮助生成潜在调节因素的假设，以供后续研究测试。选择性报告结果会导致证据不完整，可能误导决策并影响未来研究。实例在正文中以表格和文字形式描述了纳入 OMI 的测量属性结果存在不一致，但未报告具体的原因。

3.7.11 条目 27c：若进行了敏感性分析，应报告每项分析的结果，并评估主要分析的稳健性。选择性报告结果可能导致偏倚，故必须全面报告所有敏感性分析结果。最佳方式是以汇总表形式呈现所有子集的敏感性分析结果，以避免选择性报告带来的偏倚。实例未报告敏感性分析相关内容。

3.7.12 条目 28：报告每个 OMI 测量属性的证据体确定性水平，有助于读者了解整合结果的可信度。可基于确定性评估系统（如 GRADE）对证据体的确定性水平进行清晰评价和报告。建议使用证据总结表报告证据体确定性的评估结果，也可在正文中提供证据体确定性水平及其判断理由。实例在正文中以表格和文字形式报告了纳入 OMI 的测量属性汇总结果的 GRADE 评价结果。

3.7.13 条目 29：OMIs 系统评价的目的具有多样性，而作者可就 OMI 的适用性提出建议，以辅助用户进行选择。不过，有学者认为政策和实践建议超出了系统评价的范畴，而另一些学者则认为提供特定用途的适用性建议有

助于选择 OMI 并推动测量标准化。实例对纳入的评估糖尿病患者健康素养的工具给出了具体推荐。

3.8 讨论

3.8.1 条目 30a：讨论系统评价结果与其他相关证据的关系，可帮助读者理解研究发现。作者可对比当前结果与类似系统评价结果，探寻结果不一致的原因。同时，可总结系统评价未涉及但与决策相关的其他信息，如患者和临床医生偏好的证据。实例报告了其他系统评价中发现的评价测量工具时“金标准”设置不恰当等局限性。

3.8.2 条目 30b：探讨纳入研究证据的完整性、适用性与不确定性，利于读者准确解读研究结果。作者需指出可能影响结果准确性的因素，如纳入研究少、参与者少、存在偏倚风险或结果缺失等。同时，应讨论研究仅部分或间接涉及目标人群对相关性和适用性的影响。实例讨论了纳入研究的局限性，包括多数文献未详细描述研究设计及数据分析过程，重测时间设定不符 COSMIN 指南建议且未解释原因，未对量表的测量误差及反应性进行报告等。

3.8.3 条目 30c：讨论系统评价过程中存在的局限性有助于理解整合结果的可信度。作者可能报告如仅纳入英文研究、检索数据库少、单一研究者筛选记录或收集数据、未联系原始研究作者澄清模糊信息等内容，并讨论未能获取所有潜在研究报告或数据不足的影响。实例未报告其系统评价过程中的局限性。

3.8.4 条目 30d：OMI 系统评价旨在为各类用户（如研究人员、医疗提供者、患者及政策制定者等）提供决策参考。作者应探讨 OMI 适用性对实践与政策的影响，阐明不同测量属性结果的影响及对不同使用情境的潜在影响，以及如何依据主要结果做出不同决策。若不给出普遍适用的实践或政策建议，可强调将证据转化至不同情境的关键因素及可能改变测量属性结果的因素[11-12]。此外，明确提出未来研究的建议，如未研究的人群类型、评估的 OMI 或测量属性、理想的研究设计特征等，有助于指导未来研究工作。实例中推荐将 KHLS-DM 和 DNT-15 作为评估糖尿病患者健康素养的工具，同时建议未来研究者在开发工具或验证时严格遵循 COSMIN 指南进行研究。

4 讨论

近年来，OMI 系统评价作为一种新兴且快速发展的系统评价方法，基于 COSMIN 方法学，通过对某一结局变量的评价工具进行全面地检索和系统地评价，综合比较不同工具的测量属性，能够为研究者和决策者选择最佳测量工具提供证据参考[3, 7]。然而，大多数已发表的此类系统评价中缺乏对关键信息的报告，影响了终端用户对 OMI 的质量判断和最终选择[7, 11]。PRISMA 2020 报告规范虽然应用广泛，但其条目侧重于干预性系统评价内容的报告[25]，无法满足需要报告 OMI 系统评价中测量属性、效应评价和证据合成方法方面的特点。为了解决该问题，PRISMA-COSMIN for OMIs 2024 应运而生[8, 26]，该规范不仅提供了全文报告清单以及摘要报告清单，还提供了超过 100 页的解释性文件[11-12]，有望能够为 OMI 系统评价的透明和完整报告提供有价值的指导。建议有关期刊尽快将其纳入稿约，要求 OMI 系统评价作者遵循 PRISMA-COSMIN for OMIs 2024 规范撰写手稿和报告结果。本文结合一篇发表于《中华护理杂志》的实例研究对该规范进行了解读，发现了该研究报告中存在的不足；后续本团队将基于该工具全面评价中文期刊发表的此类系统评价[27]，以期能让国内研究者更好地了解国内当前该领域存在的问题，从而促进有关学者更好地理解和应用该报告规范，规范相关系统评价的撰写和报告，进而提高其整体质量[28-29]，最终为改善患者结局和促进卫生保健产生促进作用。

作者贡献：张强负责初稿撰写与经费支持；王子怡、闫乐、梁浩卫、王金钒、方进华负责论文审查及修订；赵文霞负责研究指导、论文审查及修订；卢存存负责研究选题、论文审查及修订。

本文无利益冲突。

张强 https://orcid.org/0000-0003-4849-4700

赵文霞 https://orcid.org/000-0001-6666-9469

卢存存 https://orcid.org/0000-0001-9541-9733

参考文献

[1] 邱瑞瑾，孙杨，胡嘉元，等. 临床研究中选择结局指标测量工具的方法[J]. 中国循证医学杂志，2018，18(2)：238-243. DOI：10.7507/1672-2531.201709017.

[2] 陈千吉，陈红，张英，等. 患者报告结局测量工具选择路径：以中国腰痛患者日常生活活动能力量表的选择为例[J]. 中国全科医学，2021，24(36)：4648-4652，4660. DOI：10.12114/j.issn.1007-9572.2021.02.068.

[3] PRINSEN C C，MOKKINK L B，BOUTER L M，et al. COSMIN guideline for systematic reviews of patient-reported outcome measures [J]. Qual Life Res，2018，27(5)：1147-1157. DOI：10.1007/s11136-018-1798-3.

[4] 韩舒羽，杨宪暇，张宜竹，等. COSMIN-RoB 清单中稳定性和测量误差偏倚风险清单更新内容解读[J]. 中国循证医学杂志，2022，22(10)：1212-1218.

[5] 柴芯茹，韩舒羽，张宜竹，等. 基于 COSMIN 方法的结局指标测评工具系统评价科研辅助工具框架与功能条目的构建研究[J]. 中国循证医学杂志，2023，23(11)：1313-1317.

[6] XIE X M，GAO J，BAI D X，et al. Assessment tools for stigma in breast cancer patients based on COSMIN guidelines：a systematic review[J]. Support Care Cancer，2023，32(1)：65. DOI：10.1007/s00520-023-08276-3.

[7] MOKKINK L B，TERWEE C B，STRATFORD P W，et al.



Evaluation of the methodological quality of systematic reviews of health status measurement instruments [J]. Qual Life Res，2009，18（3）：313-333. DOI：10.1007/s11136-009-9451-9.

[8] ELSMAN E B M，BUTCHER N J，MOKKINK L B，et al. Study protocol for developing，piloting and disseminating the PRISMA-COSMIN guideline：a new reporting guideline for systematic reviews of outcome measurement instruments [J]. Syst Rev，2022，11（1）：121. DOI：10.1186/s13643-022-01994-5.

[9] MCKENNA S P，HEANEY A. Setting and maintaining standards for patient-reported outcome measures：can we rely on the COSMIN checklists? [J]. J Med Econ，2021，24（1）：502-511. DOI：10.1080/13696998.2021.1907092.

[10] 高亚，刘明，杨珂璐，等. 系统评价报告规范：PRISMA 2020 与 PRISMA 2009 的对比分析与实例解读 [J]. 中国循证医学杂志，2021，21（5）：606-616. DOI：10.7507/1672-2531.202104143.

[11] ELSMAN E B M，MOKKINK L B，TERWEE C B，et al. Guideline for reporting systematic reviews of outcome measurement instruments（OMIs）：PRISMA-COSMIN for OMIs 2024 [J]. J Clin Epidemiol，2024，173：111422. DOI：10.1016/j.jclinepi.2024.111422.

[12] ELSMAN E B M，MOKKINK L B，TERWEE C B，et al. Guideline for reporting systematic reviews of outcome measurement instruments（OMIs）：PRISMA-COSMIN for OMIs 2024 [J]. Health Qual Life Outcomes，2024，22（1）：48. DOI：10.1186/s12955-024-02256-9.

[13] 刘素香，边莉，张叶群，等. 基于 COSMIN 指南对糖尿病患者健康素养评估工具的系统评价 [J]. 中华护理杂志，2023，58（21）：2670-2677. DOI：10.3761/j.issn.0254-1769.2023.21.016.

[14] PIEPER D，ROMBEY T. Where to prospectively register a systematic review [J]. Syst Rev，2022，11（1）：8. DOI：10.1186/s13643-021-01877-1.

[15] KROON F P B，VAN DER HEIJDE D，MAXWELL L J，et al. Core outcome measurement instrument selection for physical function in hand osteoarthritis using the OMERACT Filter 2.1 process [J]. Semin Arthritis Rheum，2021，51（6）：1311-1319. DOI：10.1016/j.semarthrit.2021.08.014.

[16] SAYFI S，CHARIDE R，ELLIOTT S A，et al. A multimethods randomized trial found that plain language versions improved adults understanding of health recommendations [J]. J Clin Epidemiol，2024，165：111219. DOI：10.1016/j.jclinepi.2023.11.009.

[17] SIENA L M，PAPAMANOLIS L，SIEBERT M J，et al. Industry involvement and transparency in the most cited clinical trials，2019-2022 [J]. JAMA Netw Open，2023，6（11）：e2343425. DOI：10.1001/jamanetworkopen.2023.43425.

[18] LU C C，KE L X，LI J Y，et al. Saffron（Crocus sativus L.）and health outcomes：a meta-research review of meta-analyses and an evidence mapping study [J]. Phytomedicine，2021，91：153699. DOI：10.1016/j.phymed.2021.153699.

[19] PORRITT K，GOMERSALL J，LOCKWOOD C. JBI's systematic reviews：study selection and critical appraisal [J]. Am J Nurs，2014，114（6）：47-52. DOI：10.1097/01.NAJ.0000450430.97383.64.

[20] RETHLEFSEN M L，KIRTLEY S，WAFFENSCHMIDT S，et al. PRISMA-S：an extension to the PRISMA statement for reporting literature searches in systematic reviews [J]. Syst Rev，2021，10（1）：39. DOI：10.1186/s13643-020-01542-z.

[21] MUKA T，GLISIC M，MILIC J，et al. A 24-step guide on how to design，conduct，and successfully publish a systematic review and meta-analysis in medical research [J]. Eur J Epidemiol，2020，35（1）：49-60. DOI：10.1007/s10654-019-00576-5.

[22] DEVJI T，CARRASCO-LABRA A，QASIM A，et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes：instrument development and reliability study [J]. BMJ，2020，369：m1714. DOI：10.1136/bmj.m1714.

[23] WANG Y T，DEVJI T，CARRASCO-LABRA A，et al. A step-by-step approach for selecting an optimal minimal important difference [J]. BMJ，2023，381：e073822. DOI：10.1136/bmj-2022-073822.

[24] 陈祎婷，沈蓝君，彭健，等. 改良版定量系统评价证据分级方法对患者报告结局测量工具的评价 [J]. 解放军护理杂志，2020，37（10）：57-60. DOI：10.3969/j.issn.1008-9993.2020.10.014.

[25] PAGE M J，MCKENZIE J E，BOSSUYT P M，et al. The PRISMA 2020 statement：an updated guideline for reporting systematic reviews [J]. BMJ，2021，372：n71. DOI：10.1136/bmj.n71.

[26] ELSMAN E B M，SMITH M，HOFSTETTER C，et al. A blueprint for patient and public involvement in the development of a reporting guideline for systematic reviews of outcome measurement instruments：PRISMA-COSMIN for OMIs 2024 [J]. Res Involv Engagem，2024，10（1）：33. DOI：10.1186/s40900-024-00563-5.

[27] 王子怡，卢存存，黄家艺，等. 中文期刊发表的预测模型系统评价文献调查与评价：方法学质量和报告质量 [J]. 协和医学杂志，2024，15（4）：927-935. DOI：10.12290/xhyxzz.2023-0418.

[28] HUGHES-NOEHRER L，AUBERT BONN N，DE MARIA M，et al. UK Reproducibility Network open and transparent research practices survey dataset [J]. Sci Data，2024，11（1）：912. DOI：10.1038/s41597-024-03786-z.

[29] BOAYE BELLE A，ZHAO Y X. Evidence-based decision-making：on the use of systematicity cases to check the compliance of reviews with reporting guidelines such as PRISMA 2020 [J]. Expert Syst Appl，2023，217：119569. DOI：10.1016/j.eswa.2023.119569.

（收稿日期：2024-10-15；修回日期：2025-02-12）
（本文编辑：崔莎）